

ON OPTIMALITY PROPERTIES OF THE SHIRYAEV-ROBERTS PROCEDURE

BY MOSHE POLLAK^{*} AND ALEXANDER G. TARTAKOVSKY^{†‡}

The Hebrew University of Jerusalem and the University of Southern California

We consider the simple changepoint problem setting, where observations are independent, iid pre-change and iid post-change, with known pre- and post-change distributions. The Shiryaev-Roberts detection procedure is known to be asymptotically minimax in the sense of minimizing maximal expected detection delay subject to a bound on the average run length to false alarm, as the latter goes to infinity. Here we present other optimality properties of the Shiryaev-Roberts procedure.

1. Introduction. Changepoint problems deal with detecting a change in the state of a process, where information one has about the state of affairs is in the form of observations. In the sequential setting, observations are obtained sequentially and, as long as their behavior is consistent with the initial (or target) state, one is content to let the process continue. If the state changes, then one is interested in detecting that a change is in effect, usually as soon as possible after its occurrence.

Any detection policy may give rise to false alarms. Intuitively, the desire to detect a change quickly causes one to be (relatively) trigger-happy, which will bring about many false alarms if there is no change. On the other hand, attempting to avoid false alarms too strenuously will lead to a long delay between the time of occurrence of a real change and its detection. Common operating characteristics of a sequential detection policy are ARL_{2FA} = the Average Run Length (the expected number of observations) to False

^{*}Moshe Pollak is Marcy Bogen Professor of Statistics at the Hebrew University of Jerusalem. His work was supported in part by a grant from the Israel Science Foundation, by the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem, and by the U.S. Army Research Office MURI grant W911NF-06-1-0094 at the University of Southern California.

[†]The work of Alexander Tartakovsky was supported in part by the U.S. Office of Naval Research grant N00014-06-1-0110 and the U.S. Army Research Office MURI grant W911NF-06-1-0094 at the University of Southern California.

[‡]**Corresponding author.**

AMS 2000 subject classifications: Primary 62L10; 62L15; secondary 60G40.

Keywords and phrases: Change-point Problems, CUSUM Procedures, Shiryaev-Roberts Procedures; Sequential Detection

Alarm (assuming that there is no change) and the AD2D = Average Delay to Detection (the expected delay between a real change and its detection). The gist of the changepoint problem is to produce a detection policy that (at least approximately) minimizes the AD2D subject to a bound on the ARL2FA. The constitution of a good policy depends very much on what is known about the stochastic behavior of the observations, both pre- and post-change.

Let X_1, X_2, \dots denote the series of observations, and let ν be the serial number of the first post-change observation. Let \mathbf{P}_k and \mathbf{E}_k denote probability and expectation when $\nu = k$, and let \mathbf{P}_∞ and \mathbf{E}_∞ denote the same when $\nu = \infty$ (i.e., there never is a change). A sequential change detection procedure is identified with a stopping time N on X_1, X_2, \dots , i.e., $\{N \leq n\} \in \mathcal{F}_n$, where $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ is the sigma-algebra generated by the first n observations.

In this paper, we consider the simplest setting of the problem, where the observations are independent, each having density f_0 pre-change and density f_1 post-change, where both f_0 and f_1 are known, and only the value of ν , the point of change, is unknown. (In practice, often f_0 is known. Realistically, f_1 is not known, but the simple setting yields a benchmark for the best one can hope.) In this setting, Moustakides (1986) proved that the Cusum procedure (Page, 1954) is optimal in the sense of minimizing the worst-worst case (essential supremum) expected detection delay

$$\sup_{k \geq 1} \operatorname{ess\,sup}_{\mathcal{F}_k} \mathbf{E}_k[(N - k)^+ | \mathcal{F}_k]$$

over all stopping times N for which

$$(1) \quad \text{ARL2FA}(N) = \mathbf{E}_\infty N \geq B,$$

where $B > 0$ is a value set before the surveillance begins. See also Lorden (1971) and Ritov (1990). For a continuous-time Brownian motion a similar result has been established by Beibel (1996) and Shiryaev (1996).

Pollak (1985) proved that the Shiryaev-Roberts procedure (Roberts, 1966; Shiryaev, 1963) is asymptotically (as $B \rightarrow \infty$) optimal in the sense of minimizing the supremum AD2D

$$\sup_{k \geq 1} \mathbf{E}_k(N - k | N \geq k)$$

over all stopping times N that satisfy (1).

Here we prove other (exact) optimality properties of the Shiryaev-Roberts detection procedure. To be specific, in Section 2, we prove that the Shiryaev-Roberts procedure is (exactly) optimal in the sense of minimizing the “integral AD2D” $= \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+$ for every $B > 0$ in the class of procedures

with the ARL2FA constraint (1). In Section 3, we consider the setting where a change occurs in a distant future (i.e., ν is large) and the detection of a change is preceded by a large number of false detections. We prove that the Shiryaev-Roberts procedure is the best one can do in terms of minimizing the expected detection delay asymptotically when $\nu \rightarrow \infty$ in the class (1), for every $B > 0$.

Both problem settings have been previously considered for a continuous-time Brownian motion model. See Feinberg and Shiryaev (2006); Shiryaev (1963) and Remarks 1 and 3 below.

2. Minimizing integral AD2D. Using the notation of the previous section, the Shiryaev-Roberts procedure calls for stopping and raising an alarm at

$$(2) \quad N_{AB} = \min \{n \geq 1 : R_n \geq A_B\},$$

where

$$(3) \quad R_n = \sum_{k=1}^n \frac{p(X_1, \dots, X_n | \nu = k)}{p(X_1, \dots, X_n | \nu = \infty)} = \sum_{k=1}^n \prod_{i=k}^n \frac{f_1(X_i)}{f_0(X_i)}$$

and A_B is such that $\mathbf{E}_\infty N_{AB} = B$.

Below in Theorem 1 we prove that the Shiryaev-Roberts procedure is exactly optimal in the sense of minimizing the integral $\text{AD2D} = \sum_{k=1}^\infty \mathbf{E}_k(N - k)^+$ in the class of detection procedures $\Delta_B = \{N : \text{ARL2FA}(N) \geq B\}$ in which the mean time to false alarm is not less than the given positive number B . We begin with a sketch of the argument why one may expect this to be true.

To this end, we first need to consider the following Bayesian problem, denoted by $\mathcal{B}(\rho, c)$. Suppose ν is random and has a geometric prior distribution

$$\mathbf{P}(\nu = k) = \rho(1 - \rho)^{k-1}, \quad k \geq 1,$$

and the losses associated with stopping at time N are 1 if $N < \nu$ and $c \cdot (N - \nu)$ if $N \geq \nu$, where $0 < \rho < 1$ and $c > 0$ are fixed constants. Write $\mathbf{P}^\rho(\bullet) = \sum_{k=1}^\infty \rho(1 - \rho)^{k-1} \mathbf{P}_k(\bullet)$ for the “average” probability and \mathbf{E}^ρ for the corresponding expectation.

Solution of $\mathcal{B}(\rho, c)$ requires minimization of the expected loss

$$(4) \quad \varphi_{c,\rho}(N) = \mathbf{P}^\rho(N < \nu) + c\mathbf{E}^\rho(N - \nu)^+,$$

and the Bayes rule for this problem is given by the Shiryaev procedure (cf. Shiryaev, 1963, 1978), which is the stopping time

$$(5) \quad T_{\rho,c} = \min \{n \geq 1 : \mathbf{P}^\rho(\nu \leq n | \mathcal{F}_n) \geq \delta_{\rho,c}\},$$

where $0 < \delta_{\rho,c} < 1$ is an appropriate threshold.

Obviously, the $\mathcal{B}(\rho, c)$ problem is equivalent to maximizing

$$\frac{1}{\rho}[1 - \varphi_{c,\rho}(N)] = \frac{\mathbf{P}^\rho(N \geq \nu)}{\rho} - c \frac{\mathbf{E}^\rho(N - \nu)^+}{\rho}.$$

In the proof of Theorem 1 below, we show that, for any stopping time N ,

$$\frac{\mathbf{P}^\rho(N \geq \nu)}{\rho} \xrightarrow[\rho \rightarrow 0]{} \mathbf{E}_\infty N, \quad \frac{\mathbf{E}^\rho(N - \nu)^+}{\rho} \xrightarrow[\rho \rightarrow 0]{} \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+.$$

Hence

$$\frac{1}{\rho}[1 - \varphi_{c,\rho}(N)] \xrightarrow[\rho \rightarrow 0]{} \mathbf{E}_\infty N - c \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+,$$

which should be maximized in the class Δ_B .

We also show that the Shiryaev procedure $T_{\rho,c}$ converges to the Shiryaev-Roberts procedure N_{A_B} as $\rho \rightarrow 0$. Therefore, it stands to reason that the integral $\text{AD2D} = \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+$ is minimized subject to $\mathbf{E}_\infty N \geq B$.

Formal details are given in the following theorem and its proof.

THEOREM 1. *Let A_B be chosen so that $\text{ARL2FA}(N_{A_B}) = B$. Then the Shiryaev-Roberts procedure defined by (2) and (3) minimizes*

$$(6) \quad \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+$$

over all stopping times N that satisfy $\mathbf{E}_\infty N \geq B$, i.e.,

$$\inf_{N \in \Delta_B} \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+ = \sum_{k=1}^{\infty} \mathbf{E}_k(N_{A_B} - k)^+ \quad \text{for every } B > 0,$$

where $\Delta_B = \{N : \text{ARL2FA}(N) \geq B\}$.

PROOF. Consider the Bayesian problem $\mathcal{B}(\rho, c)$ with Geometric(ρ) prior distribution and the average loss (4). Shiryaev (1963, 1978) proved that the expected loss (4) for the problem $\mathcal{B}(\rho, c)$ is minimized by the stopping time (5). Applying Bayes' formula, it is easy to see that

$$\mathbf{P}^\rho(\nu \leq n | \mathcal{F}_n) = \frac{R_{\rho,n}}{R_{\rho,n} + 1/\rho},$$

where

$$R_{\rho,n} = \sum_{k=1}^n \prod_{i=k}^n \left(\frac{1}{1 - \rho} \frac{f_1(X_i)}{f_0(X_i)} \right).$$

Hence, the Shiryaev rule can be written in the equivalent form

$$(7) \quad T_{\rho,c} = \min \{n \geq 1 : R_{\rho,n} \geq A_{\rho,c}\},$$

where $A_{\rho,c} = (1/\rho)[\delta_{\rho,c}/(1 - \delta_{\rho,c})]$.

Note first that $R_{\rho,n} \xrightarrow{\rho \rightarrow 0} R_n$.

By Theorem 1 of Pollak (1985), there exist a constant $0 < c^* < \infty$ and a sequence $\{\rho_i, c_i\}_{i=1}^{\infty}$ with $\rho_i \xrightarrow{i \rightarrow \infty} 0$, $c_i \xrightarrow{i \rightarrow \infty} c^*$ such that N_{AB} is the limit of the Bayes rules T_{ρ_i, c_i} as $i \rightarrow \infty$. Furthermore,

$$(8) \quad \limsup_{p \rightarrow 0, c \rightarrow c^*} \frac{1 - \varphi_{c,\rho}(T_{\rho,c})}{1 - \varphi_{c,\rho}(N_{AB})} = 1,$$

where $\varphi_{c,\rho}(N)$ is the expected loss associated with using the stopping time N for $\mathcal{B}(\rho, c)$.

Now, for any stopping time N ,

$$\begin{aligned} \frac{1}{\rho}[1 - \varphi_{c,\rho}(N)] &= \frac{1}{\rho} [(1 - \mathbf{P}^\rho(N < \nu)) - c\mathbf{E}^\rho(N - \nu)^+] \\ &= \frac{\mathbf{P}^\rho(N \geq \nu)}{\rho} [1 - c\mathbf{E}^\rho(N - \nu | N \geq \nu)]. \end{aligned}$$

Since

$$\begin{aligned} \frac{\mathbf{P}^\rho(N \geq \nu)}{\rho} &= \frac{1}{\rho} \sum_{k=1}^{\infty} \mathbf{P}_k(N \geq k) \rho (1 - \rho)^{k-1} \\ &= \sum_{k=1}^{\infty} \mathbf{P}_\infty(N \geq k) (1 - \rho)^{k-1} \\ &\xrightarrow{\rho \rightarrow 0} \sum_{k=1}^{\infty} \mathbf{P}_\infty(N \geq k) = \mathbf{E}_\infty N \end{aligned}$$

and

$$\begin{aligned} \frac{\mathbf{P}^\rho(N \geq \nu) \mathbf{E}^\rho(N - \nu | N \geq \nu)}{\rho} &= \frac{\mathbf{E}^\rho(N - \nu; N \geq \nu)}{\rho} \\ &= \frac{1}{\rho} \sum_{k=1}^{\infty} \mathbf{E}_k(N - k; N \geq k) \rho (1 - \rho)^{k-1} \\ &= \sum_{k=1}^{\infty} \mathbf{E}_k(N - k; N \geq k) (1 - \rho)^{k-1} \\ &\xrightarrow{\rho \rightarrow 0} \sum_{k=1}^{\infty} \mathbf{E}_k(N - k; N \geq k) = \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+, \end{aligned}$$

it follows that for any stopping time N that has finite ARL2FA

$$\frac{1}{\rho}[1 - \varphi_{c,\rho}(N)] \xrightarrow[\rho \rightarrow 0]{} \mathbf{E}_\infty N - c \sum_{k=1}^{\infty} \mathbf{E}_k(N - k)^+,$$

which together with (8) establishes that the Shiryaev-Roberts procedure minimizes (6) over all stopping times that satisfy $\mathbf{E}_\infty N = B$. Note that if $B_1 > B$, then $N_{A_{B_1}}$ is stochastically larger than N_{A_B} , i.e., all expectations in (6) become larger. This implies that the Shiryaev-Roberts procedure minimizes (6) in the class Δ_B . This completes the proof of the theorem. \square

COROLLARY 1. *The Shiryaev-Roberts procedure defined by (2) and (3) minimizes*

$$(9) \quad \frac{\sum_{k=1}^{\infty} \mathbf{E}_k(N - k | N \geq k) \mathbf{P}_\infty(N \geq k)}{\sum_{j=1}^{\infty} \mathbf{P}_\infty(N \geq j)}$$

over all stopping times N that satisfy $\mathbf{E}_\infty N = B$, i.e.,

$$\inf_{\{N: \mathbf{E}_\infty N = B\}} \sum_{k=1}^{\infty} w_k(N) \mathbf{E}_k(N - k | N \geq k) = \sum_{k=1}^{\infty} w_k(N_{A_B}) \mathbf{E}_k(N_{A_B} - k | N_{A_B} \geq k),$$

where

$$w_k(N) = \frac{\mathbf{P}_\infty(N \geq k)}{\sum_{j=1}^{\infty} \mathbf{P}_\infty(N \geq j)}$$

and the threshold A_B is selected so that $\mathbf{E}_\infty N_{A_B} = B$.

PROOF. Obviously, $\sum_{j=1}^{\infty} \mathbf{P}_\infty(N \geq j) = \mathbf{E}_\infty N = B$, so the denominator in (9) is constant over all stopping times under consideration. As for the numerator,

$$(10) \quad \begin{aligned} \mathbf{E}_k(N - k | N \geq k) \mathbf{P}_\infty(N \geq k) &= \mathbf{E}_k(N - k | N \geq k) \mathbf{P}_k(N \geq k) \\ &= \mathbf{E}_k(N - k; N \geq k) = \mathbf{E}_k(N - k)^+. \end{aligned}$$

Application of Theorem 1 concludes the proof. \square

REMARK 1. Recently, Feinberg and Shiryaev (2006) established a result similar to Theorem 1 for Brownian motion where an abrupt change occurs in the drift, in which case the integral AD2D is $\int_0^\infty \mathbf{E}_\nu(N - \nu)^+ d\nu$. They refer to this as ‘‘A Generalized Bayesian Setting.’’

While Theorem 1 and Corollary 1 are of interest in their own right, they are useful for proving other interesting optimality results, as it will become apparent in the next section.

3. Optimality for a change appearing after many re-runs. Consider a context in which it is of utmost importance to detect a real change as quickly as possible after its occurrence, even at the price of raising many false alarms (using a repeated application of the same stopping rule) before the change occurs. This essentially means that the changepoint ν is very large compared to the constant B which, in this case, defines the mean time between consecutive false alarms.

To be more specific, let $N_{AB}^{(1)}, N_{AB}^{(2)}, \dots$ be sequential independent repetitions of N_{AB} defined in (2), i.e.,

$$(11) \quad N_{AB}^{(i)} = \min \left\{ n > \sum_{j=1}^{i-1} N_{AB}^{(j)} : R_n^{(i)} \geq A_B \right\} - \sum_{j=1}^{i-1} N_{AB}^{(j)},$$

where $N_{AB}^{(0)} = 0$ and

$$(12) \quad R_n^{(i)} = \sum_{k=N_{AB}^{(i-1)}+1}^n \prod_{i=k}^n \frac{f_1(X_i)}{f_0(X_i)}, \quad \sum_{j=1}^{i-1} N_{AB}^{(j)} < n \leq \sum_{j=1}^i N_{AB}^{(j)}.$$

Therefore, $R_n^{(i)}$, $n > \sum_{j=1}^{i-1} N_{AB}^{(j)}$ is nothing but the Shiryaev-Roberts statistic that is renewed from scratch after the $(i-1)$ st false alarm (under \mathbf{P}_∞) and is applied to the segment of data

$$X_{\sum_{j=1}^{i-1} N_{AB}^{(j)}+1}, X_{\sum_{j=1}^{i-1} N_{AB}^{(j)}+2}, \dots$$

Note that $\mathbf{E}_\infty N_{AB}^{(i)} = B$ for $i \geq 1$.

Let, for $j \geq 1$,

$$(13) \quad Q_j = N_{AB}^{(1)} + N_{AB}^{(2)} + \dots + N_{AB}^{(j)}$$

be the time of the j -th alarm, and let $J_\nu = \min\{j \geq 1 : Q_j \geq \nu\}$, i.e., Q_{J_ν} is the time of detection of a true change that occurs at ν after $J_\nu - 1$ false alarms have been raised.

Our next theorem states that the Shiryaev-Roberts procedure defined by Q_{J_ν} is asymptotically (as $\nu \rightarrow \infty$) optimal with respect to the expected delay $\mathbf{E}_\nu(Q_{J_\nu} - \nu)$ in the class of detection procedures for which the mean time between false alarms is not less than B . Note that this result is not asymptotic with respect to the ARL2FA. In fact, it holds for every positive B .

THEOREM 2. Let ν be the time of the change. Let $N_{AB}^{(1)}, N_{AB}^{(2)}, \dots$ be sequential independent repetitions of N_{AB} as defined in (11) and let Q_1, Q_2, \dots be as in (13). Let $J_\nu = \min\{j : Q_j \geq \nu\}$.

(i) $\lim_{\nu \rightarrow \infty} \mathbf{E}_\nu(Q_{J_\nu} - \nu)$ exists.

(ii) Suppose a detection procedure N with $\text{ARL2FA}(N) = B$ is applied repeatedly. Let N_1, N_2, \dots be sequential repetitions of N , let $W_j = \sum_{i=1}^j N_i$, and let $K_\nu = \min\{j : W_j \geq \nu\}$. Then, for every $B > 0$,

$$(14) \quad \lim_{\nu \rightarrow \infty} \mathbf{E}_\nu(Q_{J_\nu} - \nu) \leq \lim_{\nu \rightarrow \infty} \mathbf{E}_\nu(W_{K_\nu} - \nu).$$

(iii) Inequality (14) holds for all $N \in \Delta_B$, where $\Delta_B = \{N : \mathbf{E}_\infty N \geq B\}$.

PROOF. Proof of (i). By renewal theory, the distribution of $\nu - Q_{J_\nu-1}$ has a limit

$$(15) \quad \lim_{\nu \rightarrow \infty} \mathbf{P}_\nu(\nu - Q_{J_\nu-1} = k) = \frac{\mathbf{P}_\infty(N_{AB} \geq k)}{\sum_{j=1}^{\infty} \mathbf{P}_\infty(N_{AB} \geq j)}$$

(see, e.g., Feller, 1966, page 356).

Using (15) and letting N_{AB} be independent of $N_{AB}^{(1)}, N_{AB}^{(2)}, \dots$, we obtain

$$\begin{aligned} \mathbf{E}_\nu(Q_{J_\nu} - \nu) &= \mathbf{E}_\nu \mathbf{E}_\nu(Q_{J_\nu} - \nu | Q_{J_\nu-1}) \\ &= \sum_{k=1}^{\nu} \mathbf{E}_k(N_{AB} - k | \nu - Q_{J_\nu-1} = k, N_{AB} \geq k) \mathbf{P}_\infty(\nu - Q_{J_\nu-1} = k) \\ &= \sum_{k=1}^{\nu} \mathbf{E}_k(N_{AB} - k | N_{AB} \geq k) \mathbf{P}_\infty(\nu - Q_{J_\nu-1} = k) \\ &\xrightarrow{\nu \rightarrow \infty} \frac{\sum_{k=1}^{\infty} \mathbf{E}_k(N_{AB} - k | N_{AB} \geq k) \mathbf{P}_\infty(N_{AB} \geq k)}{\sum_{j=1}^{\infty} \mathbf{P}_\infty(N_{AB} \geq j)} \\ &= \frac{\sum_{k=1}^{\infty} \mathbf{E}_k(N_{AB} - k)^+}{\mathbf{E}_\infty N_{AB}} = \frac{\sum_{k=1}^{\infty} \mathbf{E}_k(N_{AB} - k)^+}{B}, \end{aligned}$$

which completes the proof of (i).

Proof of (ii). The same argument as in the proof of (i) yields

$$\lim_{\nu \rightarrow \infty} \mathbf{E}_\nu(W_{K_\nu} - \nu) = \frac{\sum_{k=1}^{\infty} \mathbf{E}_k(W - k)^+}{B}.$$

Combining this with Corollary 1 concludes the proof.

Proof of (iii). Write $\text{AD2D}(B) = \lim_{\nu \rightarrow \infty} \mathbf{E}_\nu(Q_{J_\nu} - \nu)$ for the AD2D of the Shiryaev-Roberts procedure N_{AB} . Note that $\text{AD2D}(B)$ tends to 0 as $B \rightarrow 0$

and to ∞ as $B \rightarrow \infty$. By virtue of (ii), it suffices to show that $\text{AD2D}(B)$ is nondecreasing in B .

Note that $\text{AD2D}(B)$ is continuous in B . Therefore, if $\text{AD2D}(B)$ were not nondecreasing in B , there would exist $0 < B_1 < B_2 < \infty$ such that $\text{AD2D}(B_1) = \text{AD2D}(B_2)$ and $\text{AD2D}(B) > \text{AD2D}(B_1) = \text{AD2D}(B_2)$ for all $B_1 < B < B_2$.

Consider the following renewal-theoretic argument. Let L_1, L_2, \dots and M_1, M_2, \dots be independent sequences of positive random variables having finite means, each of them iid. Let G^L be the asymptotic distribution of the residual waiting time of the sequence $\{L_i\}$ (i.e., of the overshoot of the sequence $\{\sum_{i=1}^j L_i\}_{j=1}^\infty$ over t , as $t \rightarrow \infty$) and let G^M be that of the sequence $\{M_i\}$. Let G^T be the asymptotic distribution of the residual waiting time for the sequence $\{T_i\}$ that is defined as follows: $\mathbf{P}(T_i = L_i) = \mathbf{P}(T_i = M_i) = 1/2$. By the usual renewal-theoretic apparatus, one can show that

$$(a) \quad G^T = \frac{\mathbf{E}L}{\mathbf{E}L + \mathbf{E}M} G^L + \frac{\mathbf{E}M}{\mathbf{E}L + \mathbf{E}M} G^M.$$

(b) Let $n_t = \min\{n : \sum_{i=1}^n T_i \geq t\}$. The asymptotic probability (as $t \rightarrow \infty$) that T_{n_t} is of the type L, M is $\mathbf{E}L/(\mathbf{E}L + \mathbf{E}M)$, $\mathbf{E}M/(\mathbf{E}L + \mathbf{E}M)$, respectively.

(c) Conditional on T_{n_t} being the type L, M the asymptotic (as $t \rightarrow \infty$) distribution of the residual waiting time $\sum_{i=1}^{n_t} T_i - t$ is G^L, G^M , respectively:

$$\lim_{t \rightarrow \infty} \mathbf{P} \left(\sum_{i=1}^{n_t} T_i - t \leq x | T_{n_t} \text{ is of type } L, M \right) = G^L, G^M.$$

Now, let $L = N_{A_{B_1}}$ and $M = N_{A_{B_2}}$. Note that the notation n_t in terms of L and M is the same as J_t in terms of $N_{A_{B_1}}$ and $N_{A_{B_2}}$. Recall that the procedure based on T “recycles” every time the Shiryayev-Roberts statistic crosses the boundary (A_{B_1} if the cycle has T of type $N_{A_{B_1}}$ and A_{B_2} if the cycle has T of type $N_{A_{B_2}}$). Let R_t^N be the value of the detection statistic at time t , where N is a generic stopping time that is applied repeatedly.

Let $R_n^{(i)}(N_{A_{B_1}})$ be equal to $R_n^{(i)}$ of (12) for $B = B_1$ and let $R_n^{(i)}(N_{A_{B_2}})$ be the same for $B = B_2$. To emphasize the dependence of J_ν and Q_j on the stopping time N being used, we will write J_ν^N and $Q^N(j)$. With this notation, for $j = 1, 2$,

$$\begin{aligned} & \mathbf{P}_\nu \left(R_\nu^T \leq x | Q^T(J_\nu^T - 1) = \nu - k, T_{J_\nu^T} \text{ is of type } N_{A_{B_j}} \right) \\ &= \mathbf{P}_\infty \left(R_k^{(1)}(N_{A_{B_j}}) \leq x | N_{A_{B_j}} \geq k \right) \stackrel{\text{def}}{=} F_{j,k}(x). \end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbf{P}_\nu \left(R_\nu^T \leq x \right) \\
&= \frac{\mathbf{E}_\infty N_{A_{B_1}}}{\mathbf{E}_\infty N_{A_{B_1}} + \mathbf{E}_\infty N_{A_{B_2}}} \sum_{k=1}^{\infty} F_{1,k}(x) \\
&\quad \times \mathbf{P}_\infty \left(Q^{N_{A_{B_1}}} (J_\nu^{N_{A_{B_1}}}) = \nu - k \mid T_{J_\nu^T} \text{ is of type } N_{A_{B_1}} \right) \\
&+ \frac{\mathbf{E}_\infty N_{A_{B_2}}}{\mathbf{E}_\infty N_{A_{B_1}} + \mathbf{E}_\infty N_{A_{B_2}}} \sum_{k=1}^{\infty} F_{2,k}(x) \\
&\quad \times \mathbf{P}_\infty \left(Q^{N_{A_{B_2}}} (J_\nu^{N_{A_{B_2}}}) = \nu - k \mid T_{J_\nu^T} \text{ is of type } N_{A_{B_2}} \right)
\end{aligned}$$

and, by (15),

$$\begin{aligned}
& \lim_{\nu \rightarrow \infty} \mathbf{P}_\nu \left(R_\nu^T \leq x \right) \\
&= \frac{\mathbf{E}_\infty N_{A_{B_1}}}{\mathbf{E}_\infty N_{A_{B_1}} + \mathbf{E}_\infty N_{A_{B_2}}} \sum_{k=1}^{\infty} F_{1,k}(x) \frac{\mathbf{P}_\infty \left(N_{A_{B_1}} \geq k \right)}{\mathbf{E}_\infty N_{A_{B_1}}} \\
&\quad + \frac{\mathbf{E}_\infty N_{A_{B_2}}}{\mathbf{E}_\infty N_{A_{B_1}} + \mathbf{E}_\infty N_{A_{B_2}}} \sum_{k=1}^{\infty} F_{2,k}(x) \frac{\mathbf{P}_\infty \left(N_{A_{B_2}} \geq k \right)}{\mathbf{E}_\infty N_{A_{B_2}}}.
\end{aligned}$$

By abuse of notation, write $\text{AD2D}(N)$ for the limit (as $\nu \rightarrow \infty$) of the average delay to detection when a stopping time N is applied repeatedly. It now follows that

$$\begin{aligned}
\text{AD2D}(T) &= \frac{\mathbf{E}_\infty N_{A_{B_1}}}{\mathbf{E}_\infty N_{A_{B_1}} + \mathbf{E}_\infty N_{A_{B_2}}} \text{AD2D}(N_{A_{B_1}}) \\
&\quad + \frac{\mathbf{E}_\infty N_{A_{B_2}}}{\mathbf{E}_\infty N_{A_{B_1}} + \mathbf{E}_\infty N_{A_{B_2}}} \text{AD2D}(N_{A_{B_2}}) \\
&= \frac{B_1}{B_1 + B_2} \text{AD2D}(B_1) + \frac{B_2}{B_1 + B_2} \text{AD2D}(B_2) \\
&= \text{AD2D}(B_1) = \text{AD2D}(B_2).
\end{aligned}$$

Note that $\mathbf{E}_\infty T = \frac{1}{2} \mathbf{E}_\infty N_{A_{B_1}} + \frac{1}{2} \mathbf{E}_\infty N_{A_{B_2}} = (B_1 + B_2)/2$. By definition of B_1 and B_2 , it follows that

$$\text{AD2D}(T) < \text{AD2D}(N_{A_B}) = \text{AD2D}(B) \quad \text{for } B = \frac{1}{2}(B_1 + B_2),$$

which contradicts (ii) for $B = \frac{1}{2}(B_1 + B_2)$. \square

REMARK 2. Theorem 2(iii) implies that Corollary 1 holds for all stopping times $N \in \Delta_B$.

REMARK 3. Shiryaev (1963) proved a result similar to Theorem 2(ii) for Brownian motion when a change occurs in the drift, and called this problem “Quickest Detection of a Disorder in a Stationary Regime.”

REMARK 4. It is worth noting that Theorem 2 is important in a variety of surveillance applications such as target detection and tracking, rapid detection of intrusions in computer networks, and environmental monitoring, to name a few. In all of these applications, it is of utmost importance to detect very rapidly changes that may occur in a distant future, in which case the true detection of a real change may be preceded by a long interval with frequent false alarms that are being filtered by a separate mechanism or algorithm. For example, falsely initiated target tracks are usually filtered by a track confirmation/deletion algorithm; false detections of attacks in computer networks in anomaly-based Intrusion Detection Systems (IDS) may be filtered by Signature-based IDS algorithms, etc. See, e.g., Tartakovsky (1991); Tartakovsky et al. (2006); Tartakovsky and Veeravalli (2004). The practical implication of Theorem 2 is that in these circumstances one has reason to prefer the Shiryaev-Roberts procedure to other surveillance schemes.

References.

- BEIBEL, M. (1996). A note on Ritov’s Bayes approach to the minimax property of the Cusum procedure. *Ann. Statist.* **24** 1804–1812.
- FEINBERG, E.A. AND SHIRYAEV, A.N. (2006). Quickest detection of drift change for Brownian motion in generalized Bayesian and minimax settings, *Statistics & Decisions* **24**, Issue 4, 445–470.
- FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*. Vol. II, John Wiley & Sons, Inc., New York.
- LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42** 1987–1908.
- MOUSTAKIDES, G.V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387.
- PAGE, E.S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.
- POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227.
- RITOV, Y. (1990). Decision theoretic optimality of the Cusum procedure. *Ann. Statist.* **18** 1466–1469.
- ROBERTS, S.W. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430.
- SHIRYAEV, A.N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8** 22–46.
- SHIRYAEV, A.N. (1978). *Optimal Stopping Rules*. Springer-Verlag, New York.

- SHIRYAEV, A.N. (1996). Minimax optimality of the method of cumulative sum (cusum) in the case of continuous time. *Russian Math. Surveys* **51**, no. 4, 750–751.
- TARTAKOVSKY, A.G. (1991). *Sequential Methods in the Theory of Information Systems*. Radio i Svyaz', Moscow (In Russian).
- TARTAKOVSKY, A.G., ROZOVSKII, B.L., BLAŽEK, R., AND KIM, H. (2006). Detection of intrusions in information systems by sequential change-point methods. *Statistical Methodology* **3** Issue 3 252–340.
- TARTAKOVSKY, A.G. AND VEERAVALLI, V.V. (2004). Change-point detection in multichannel and distributed systems with applications. In: *Applications of Sequential Methodologies* (N. Mukhopadhyay, S. Datta and S. Chattopadhyay, Eds), Marcel Dekker, Inc., New York, 331–363.

HEBREW UNIVERSITY OF JERUSALEM
DEPARTMENT OF STATISTICS
MOUNT SCOPUS
JERUSALEM 91905, ISRAEL
E-MAIL: msmtp@mscc.huji.ac.il

UNIVERSITY OF SOUTHERN CALIFORNIA
DEPARTMENT OF MATHEMATICS
3620 S. VERMONT AVE
LOS ANGELES, CA 90089-2532, USA
E-MAIL: tartakov@usc.edu